Biologically motivated computationally intensive approaches to image pattern recognition

Petkov, Nikolay

Published in:
Future generation computer systems

Publication date:
1995

ELSEVIER

Future Generation Computer Systems 11 (1995) 451–465

FGCS
FUTURE GENERATION COMPUTER SYSTEMS

Biologically motivated computationally intensive approaches to image pattern recognition

Nikolay Petkov *

Centre for High Performance Computing and Institute of Mathematics and Computer Science, University of Groningen, P.O.Box 800, 9700 AV Groningen, The Netherlands

Abstract

This paper presents some of the research activities of the research group in vision as a grand challenge problem whose solution is estimated to need the power of Tflop/s computers and for which computational methods have yet to be developed. The concerned approaches are biologically motivated, in that we try to mimic and use mechanisms employed by natural vision systems, more specifically the visual system of primates. Visual information representations which are motivated by the function of the primary visual cortex, more specifically by the function of so-called simple cells, are computed. Three different methods for using such representations to solve image pattern recognition problems are presented. These are: (*i*) extraction and comparison of lower-dimension representations, (*ii*) computing optimal mappings of an image onto other images by optic flow techniques and (*iii*) application of a self-organising neural network classifier. The problems of automatic recognition and classification of visual patterns, in particular the discrimination of human faces, are used to test the usefulness and feasibility of these approaches.

Keywords: Vision; Primary visual cortex; Simple cells; 2D Gabor representations; Lower-dimension representations; Optic flow; Multi-scale matching; Artificial neural networks; Self-organisation; Kohonen networks; Parallel computing

1. Introduction

Vision is considered to be a computational problem whose (real-time) solution would require parallel computers with performance in the order of at least 10^{12} operations per second (for an estimation of the computational demands see [27,28]). This and other problems with similar computational requirements have been given the name of *grand challenge problems* [10]. In contrast to other grand challenge problems, such as climate modelling, fluid turbulence, pollution dispersion, human genome, ocean circulation, quantum chromodynamics, semiconductor modelling, combustion systems, etc., in which physical models and mathematical solution methods are available, in vision the involved research community of neurophysiologists, biophysicists and computer scientists just begins to understand some of the underlying principles. Therefore, vision presents a *double* grand challenge, in that physical and

* Email: petkov@cs.rug.nl

SSDI 0167-739X(95)00015-1

biological models and computational solution methods, next to (and actually in advance of) Tflop/s computers are needed.

This paper gives an overview of some of the activities of my research group for the development of biologically motivated computationally intensive methods for image pattern recognition. The concerned methods are biologically motivated, in that we try to mimic and use mechanisms employed by biological vision systems, more specifically the visual system of primates. Since natural selection has already provided adequate solutions to vision problems in man's natural environment, we believe that such an approach may lead to reasonable solutions to technical problems.

Extensive neurophysiological studies in the past fifty years have led to the accumulation of considerable knowledge of the visual system of primates (more precisely, such data stems primarily from experiments with monkeys and cats). The functional description of classes of visual neurons makes a substantial part of this knowledge. The discoveries of the centre-surround organisation of the receptive fields of retinal ganglion cells [9] and of the orientation selectivity of the majority of neurons in the primary visual cortex [6,7] are major milestones in the development of this area. On the basis of the results of neurophysiological research one can set up computational models of visual neurons aiming at a precise quantitative description of their functional behaviour. For instance, the concept of a receptive field function which specifies the response of a visual neuron to a light spot stimulus as a function of position has successfully been applied to describe the behaviour of retinal ganglion cells and simple primary cortical cells, enabling one to predict their responses for arbitrary visual stimuli. In Section 2 of this paper, a computational model of simple cortical cells is presented. This model is used in computer simulations in which the activities of sets of simple cells are computed. The computed quantities are visualised in subsets, referred to as cortical images, giving an opportunity for improved insights into the function of the visual system [28].

Beside these fundamental aspects aiming to extend the understanding of biological vision, researchers also try to use such computer models and simulations for building artificial vision systems to solve technical problems [2,5,16,18]. A major problem which arises here is that neurophysiological data which allows to set up computational models is available only for the early stages of the visual system (of cats and monkeys), such as the retina, the lateral geniculate nuclei and the primary visual cortex. Similar to the work of other researchers, the computer simulations we do are concerned mainly with the last of these stages, the primary visual cortex. Unfortunately, at present there is no clear understanding of how the activities of primary cortical neurons are used in the higher cortical structures. For instance, while at present it is well understood that the activation of a single simple or complex cell signals the presence of an edge or a line of particular orientation, size and position, the way in which the joint activities of many such cells are used by higher cortical structures to detect more complex visual patterns is still not known. To use the words of D.H. Hubel in his 1981 Nobel Price lecture [7], 'If our perception of a certain line or curve depends on simple or complex cells, it presumably depends on a whole set of them, and how the information from such sets of cells is assembled at subsequent stages in the path to build up what we call percept of lines or curves (if indeed anything like that happens at all) is still a complete mystery.'

In Section 3 of this paper three different ways of using the results of primary cortex simulations for solving vision tasks are sketched. The proposed methods are not claimed to have biological relevance; they should rather be considered as attempts to make use of computable intermediate representations in natural vision systems. The problems of automatic classification of visual patterns, in particular the discrimination of human faces, and automatic person identification by a face image are used to test the usefulness and feasibility of the proposed methods. In Section 4 the computational requirements and some implementation issues are addressed, conclusions are drawn and directions for future research are outlined.

2. Cortical filters and images

2.1. Computational model of visual neurons with linear spatial summation

The *receptive field* of a visual neuron is the part of the visual field within which a stimulus can influence the response of the concerned neuron. For several classes of visual neurons, such as the retinal ganglion cells, the neurons in the lateral geniculate nuclei and the simple cells in the primary visual cortex, it has been shown that the different points within the receptive field of a neuron contribute to its response via a weighted sum — neurophysiologist refer to this property as 'linear spatial summation' — followed by a half-wave rectification or, more generally, thresholding. For such neurons it is useful to introduce a so-called *receptive field function* $g(x,y)$, $(x,y) \in \Omega$ (Ω-the spatial angle domain of the visual field), which specifies the weight with which the illuminance of a point has to be taken as a function of position. The response r of such a neuron to an input image $s(x,y)$, $(x,y) \in \Omega$, can be modelled as follows:

(i) *Linear spatial summation*. An integral

$$\tilde{s} = \iint_{\Omega} s(x,y)g(x,y)\mathrm{d}x\mathrm{d}y \tag{1}$$

is evaluated in the same way as if the receptive field function $g(x,y)$ were the impulse response of a linear system.

(ii) *Thresholding and non-linear contrast normalisation*. The result $\tilde{s}$ is submitted to thresholding

$$r = 0 \text{ if } \tilde{s} \leq T, \tag{2}$$

where T is a threshold value (half-way rectification is a special case for $T = 0$), and non-linear local contrast normalisation:

$$r = \frac{r_{\infty}\tilde{s}}{c_{1/2}L + \tilde{s}} \text{ if } \tilde{s} > T, \tag{3}$$

where L is the mean illuminance within the receptive field of the concerned neuron and r_{∞} and $c_{1/2}$ are constants which specify the saturation response and the value $(\tilde{s}_{1/2}/L)$ of the contrast for which half-saturation is reached, respectively (see [8] for more details of this constants in the case of retinal ganglion cells).

2.2. Simple cells

In this study we are concerned with computer simulations of so-called *simple cells* in the primary visual cortex [6,7]. The receptive fields of such cells were found to consists of a number of oriented altering parallel excitatory and inhibitory zones (Fig. 1) and various mathematical functions have been proposed to model their linear spatial summation properties. J.G. Daugman [1] proposed to model the linear spatial summation properties of simple cells by complex two-dimensional Gabor functions which minimise the product of their variances in the space and spatial frequency domains. It has been demonstrated by Jones and Palmer that these functions adequately fit experimental data [11]. Similar to previous work [26–28], here we use this family of functions with the following slightly modified parametrisation:

$$g_{\xi,\eta,\sigma,\gamma,\theta,\lambda,\varphi}(x,y) = e^{-(x'^2+\gamma^2y'^2)/2\sigma^2}\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \tag{4}$$

$$x' = (x-\xi)\cos\theta - (y-\eta)\sin\theta$$

$$y' = (x-\xi)\sin\theta + (y-\eta)\cos\theta$$

where the arguments x and y specify the position of a point in the visual field and ξ, η, σ, γ, θ, λ and φ are parameters whose effect on the function g is next explained in more detail (Fig. 1).

The pair (ξ,η), which has the same domain Ω as the pair (x,y), specifies the *center of a receptive field* within the visual field. The parameter σ determines the (linear) *size of the receptive field*. The response of a cell to a light spot in position (x,y) which is at a distance greater than 2σ from the center (ξ,η) of the receptive field can practically be neglected; this dependence is well modelled by the Gaussian factor $e^{-(x^2+\gamma^2y'^2)/2\sigma^2}$. Neurophysiological research has shown that, on the population of all simple cells, the receptive field sizes vary considerably with the diameters of

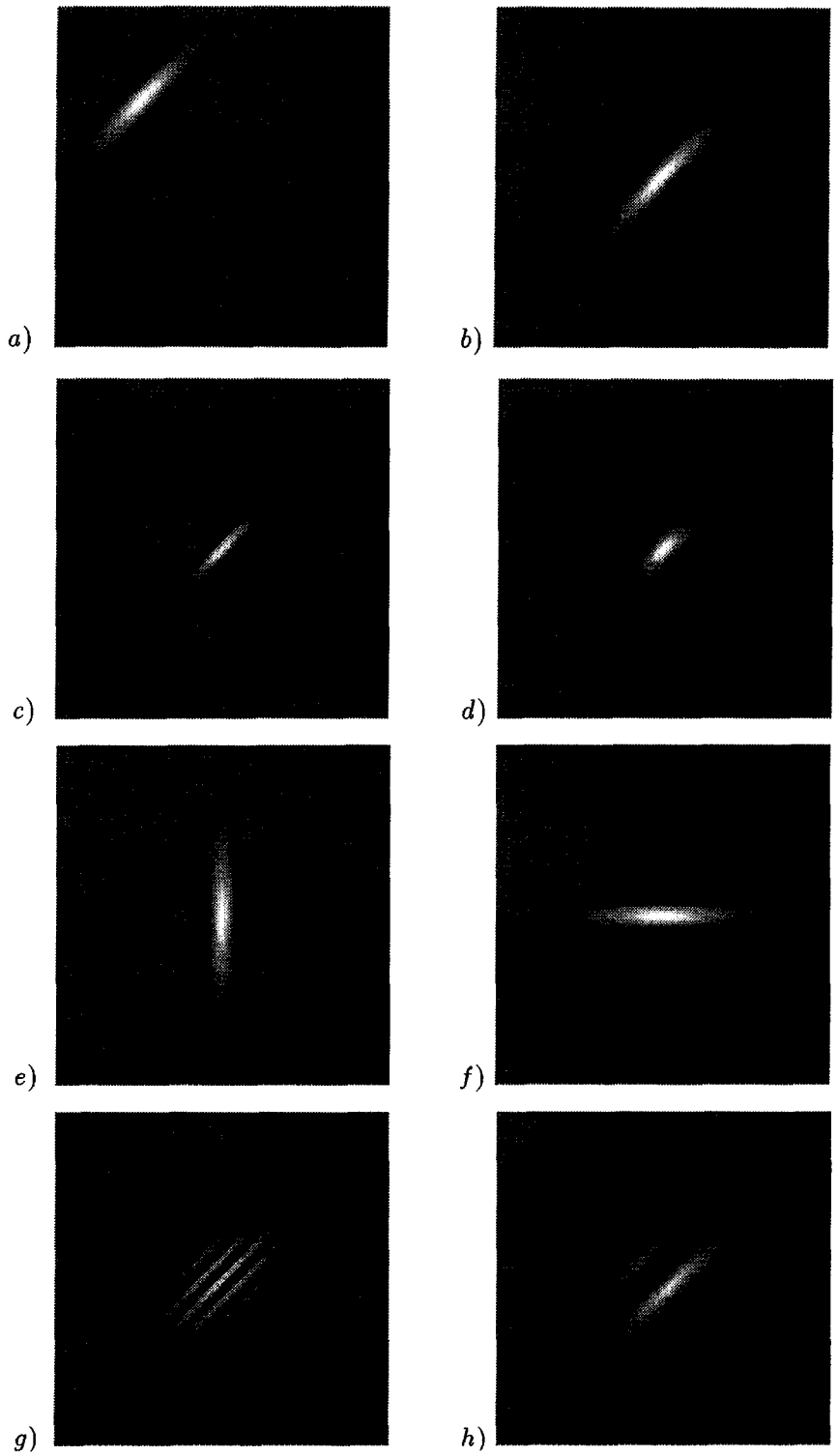

Fig. 1. Receptive fields of different positions (*a,b*), sizes (*b,c*), eccentricities (*b,d*), orientations (*e,f*), number of excitatory and inhibitory zones (*b,g*), and symmetries (*b,h*). The gray level of the background labels areas which have no effect on the response of a cell. Lighter or darker colours indicate excitatory and inhibitory zones, respectively.

the smallest to the largest receptive fields being in a ratio of at least 1:30 [1]. The eccentricity of the Gaussian factor and herewith the eccentricity of the receptive field ellipse is determined by the parameter γ, called the spatial aspect ratio. It has been found to vary in a very limited range of $0.23 < \gamma < 0.92$ [11]. The value $\gamma = 0.5$ is used in our simulations.

The angle parameter θ ($\theta \in [0,\pi)$) specifies the *orientation* of the normal to the parallel excitatory and inhibitory stripe zones (this normal is the axis x' in Eq. 4) and the parameter λ is the *wavelength* of the harmonic factor $\cos(2\pi(x'/\lambda) + \varphi)$. The ratio σ/λ determines the number of parallel excitatory and inhibitory zones which can be observed in a receptive field. Neurophysiological research shows that the parameters λ and σ are closely correlated; on the set of all cells, the ratio σ/λ which determines the spatial-frequency bandwidth of a cell varies in a very limited range of 0.4–0.9 which corresponds to two to five excitatory and inhibitory stripe zones in a receptive field [1]. The value $\sigma/\lambda = 0.5$ is used in our simulations. Typical receptive fields for the parameter values $\gamma = 0.5$ and $\sigma/\lambda = 0.5$ are shown in Fig. 1(a–c,e,f,h).

Finally, the parameter φ ($\varphi \in (-\pi,\pi]$), which is a phase offset in the argument of the harmonic factor $\cos(2\pi(x'/\lambda) + \varphi)$, determines the symmetry of the function g: for $\varphi = 0$ (Fig. 1(a–f)) and $\varphi = \pi$ (this case can be illustrated by the negatives of Fig. 1(a–f)), the function g is symmetric, or even, with respect to the center (ξ,η) of a receptive field; for $\varphi = -\frac{1}{2}\pi$ and $\varphi = \frac{1}{2}\pi$, the function g is antisymmetric, or odd (Fig. 1(h)), and all other cases are mixtures of these two. In our simulations, we use for φ the following values: 0 (symmetric receptive fields to which we refer as 'center-on' in analogy with retinal ganglion cell receptive fields whose central areas are excitatory), π (symmetric receptive fields to which we refer to as 'centre-off', since their central lobes are inhibitory) and $-\frac{1}{2}\pi$ and $\frac{1}{2}\pi$ (antisymmetric receptive fields with opposite polarity).

As to the importance of these cells for the visual system, it is believed that they play an important role in the process of form perception, in that they act as detectors of oriented intensity transitions such as edges and bars. More specifically, a cell with a symmetric receptive field function will react strongly to a bar which coincides in direction, width and polarity with the central region of the cell receptive field. A cell with an antisymmetric receptive field function will react strongly to an edge of the same orientation if the excitatory lobe is on the light side of the transition and the inhibitory lobe on its dark side.

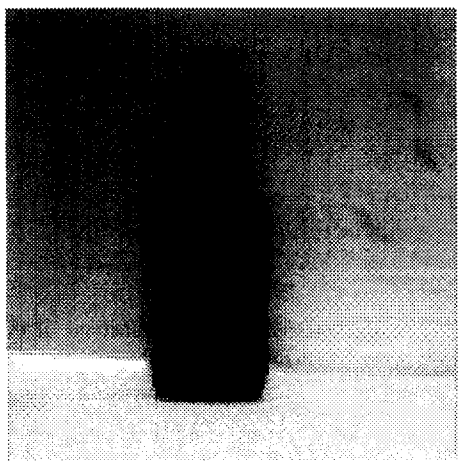

Fig. 2. An input image.

2.3. Cortical images

Substituting a receptive field function $g_{\xi,\eta,\sigma,\gamma,\theta,\lambda,\varphi}(x,y)$ in Eqs. 1–3, one can compute the response $r_{\xi,\eta,\sigma,\gamma,\theta,\lambda,\varphi}$ of a simple visual cortical cell modelled by this function to an input image $s(x,y)$. Fig. 2 shows an input image for which a number of such quantities are computed and shown in Fig. 3, grouped together in so-called *cortical images*. The quantities grouped in one such image are computed using receptive field functions with the same set os values for all parameters but the pair (ξ,η); the later pair specifies the pixel to which the value of such a quantity is assigned. (Note that the function g depends on x, y, ξ and η via the differences $x - \xi$ and $y - \eta$, so that the integral in Eq. 1 can be considered as a two-dimensional convolution of the functions s and g.) The filters which generate such images are referred to as *cortical filters or*

channels. The parameters which characterise such a filter are the orientation Θ, the size σ and the symmetry as specified by the phase φ (the parameter γ is fixed, $\gamma = 0.5$, as well as the ratio σ/λ, $\sigma/\lambda = 0.5$). The cortical images shown in Fig. 3 are generated by filters with the same value of σ (the same scale) but with different values of the parameters Θ and φ.

Roughly speaking, the effect of such a filter is to enhance illuminance transitions of a given orientation and at a give scale. The cortical filter which produced the image shown in Fig. 3(a), for instance, enhances vertical edges with a dark region on the left and a light region on the right side of such an edge. The filter which output the image shown in Fig. 3(b) also enhances vertical

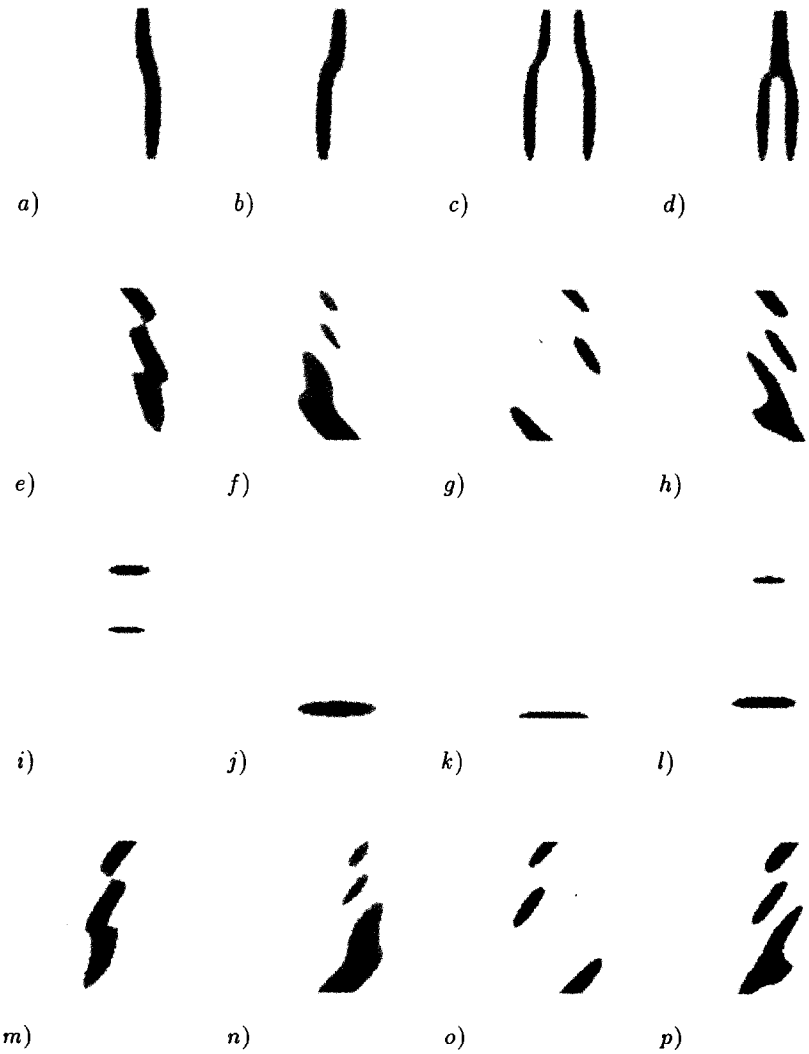

Fig. 3. Cortical images computed from the bottle input image shown in Fig. 2 applying receptive field functions of different orientations and symmetries. The images shown in one row are computed with the same value of the orientation parameter Θ as follows: $(a–d)$ $\Theta = 0$, $(e–h)$ $\Theta = \frac{1}{4}\pi$, $(i–l)$ $\Theta = \frac{1}{2}\pi$, $(m–p)$ $\frac{3}{4}\pi$. The values of the phase/symmetry parameter φ for the images in the first and second column, which represent the responses of cells with antisymmetric receptive fields of opposite polarity, are $-\frac{1}{2}\pi$ and $\frac{1}{2}\pi$, respectively. The values of φ for the images in the third and fourth column, which represent the responses of cells with symmetric centre-on and centre-off receptive fields, are 0 and π, respectively. The values of the other parameters are as follows: $\sigma = \frac{1}{16}S$ (S-size of the image side), $\gamma = 0.5$, $\lambda = 2\sigma$.

edges but with an inverse polarity of the light-dark transition. The other images in the first two columns of Fig. 3 are generated by similar edge enhancing filters which have different preferred orientations. The effect of such a filter is easy to understand: The receptive field functions used for spatial summation are antisymmetric and (for the biologically relevant values of the parameter γ and the parameter ratio σ/λ) can roughly be considered as consisting of one excitatory and one inhibitory lobe (Fig. 1(h)). When the excitatory lobe happens to be on the light side of an edge transition of appropriate orientation and the inhibitory lobe on its dark side, the result is a strong response of the filter in the corresponding pixel at the concerned edge transition.

In a similar way, one can easily infer that the main effect — and possibly the biological role — of the filters which produced the cortical images shown in the third and fourth column of Fig. 3 is to enhance lines and bars of appropriate orientation, width and polarity. Indeed, the receptive field function which determines the spatial summation properties of such a filter consists either

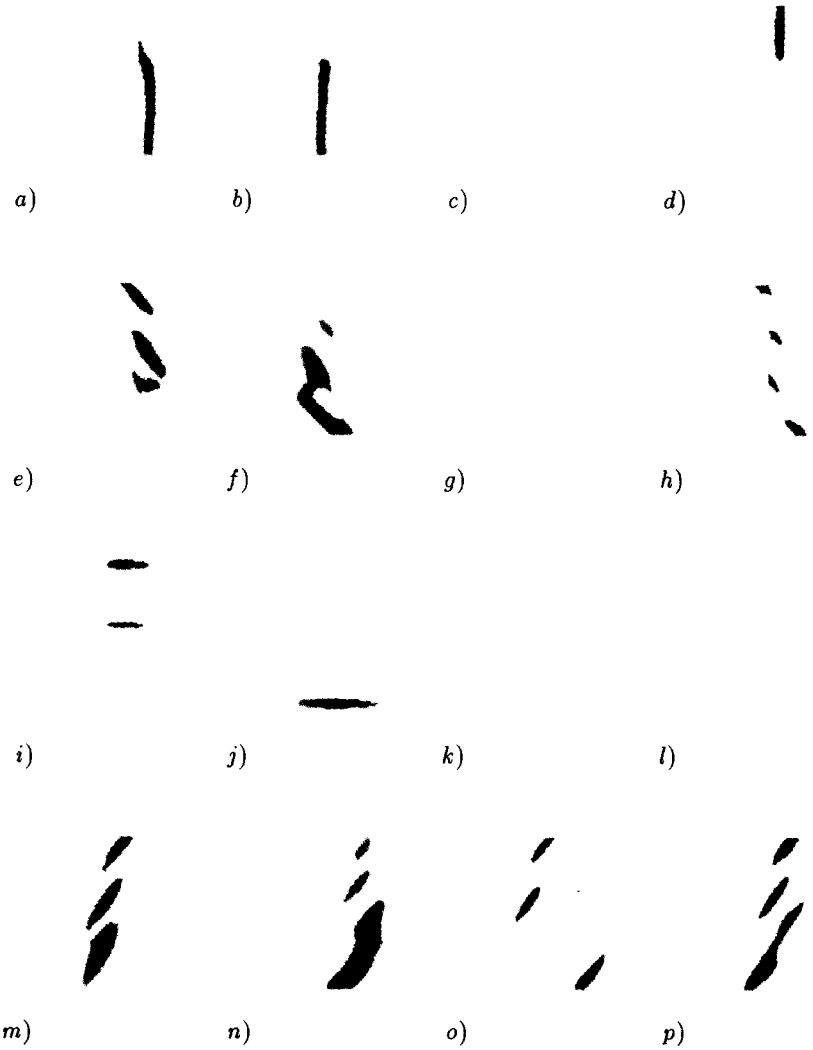

Fig. 4. Cortical images computed from the cortical images shown in Fig. 3 by lateral inhibition of the cortical images which have the same value of the orientation parameter θ (these are the images shown in one row). The resulting images are sparser and less ambiguous than the corresponding images of Fig. 3 (compare Fig. 4(a–d) with Fig. 3(a–d) and Fig. 4(i–l) with Fig. 3(i–l)).

of one excitatory central lobe enclosed by two inhibitory side lobes in the case of centre-on cells (see e.g. Fig. 1(e)), or of one inhibitory central lobe enclosed by two excitatory side lobes corresponding to centre-off cells (this would be the negative of Fig. 1(e)). When the central lobe happens to coincide in position, orientation and width with a bar of the same illuminance polarity in the input image, the result will be a strong response of the filter in the corresponding pixel position. This is quite well illustrated by Fig. 3(d) where the dark neck of the bottle is 'seen' as a bar through a filter with a vertically oriented symmetric centre-off receptive field function.

It is, of course, not so that the cortical filters with antisymmetric and symmetric receptive field functions enhance only edges or only bars, respectively. For instance, a bar which gives rise to activity in the output of a filter with a symmetric spatial summation function will also give rise to activity in the output of a filter with an antisymmetric receptive field function, since a bar has two edges. This is illustrated by Fig. 3(a,b,d): the neck of the bottle which is seen as a bar in Fig. 3(d) gives also rise to activity in the edge-detecting filters (Fig. 3(a,b)). Similarly, a filter with a symmetric spatial summation function will enhance not only bars but edges as well, an effect well illustrated by the activity at the edges of the body of the bottle in the same figures. In this way the effect of the concerned filters is broader than the effect suggested by the simple qualitative analysis given above.

For technical applications one might wish to have filters with a more distinct effect, in that such a filter should enhance edges or bars, but not both at the same time. This would simplify the interpretation of the resulting cortical images and the subsequent processing steps. Elsewhere [17,25] we proposed a mechanism for the elimination of so-called 'shadow' lines in cortical images obtained from filters with antisymmetric spatial summation functions. An edge which is strongly enhanced by a cortical filter with an antisymmetric receptive field function of orientation θ and phase φ ($\varphi = 0$ or $\varphi = \pi$) will give rise to a pair of parallel lines in the cortical image produced by a cortical filter with an antisymmetric receptive field function of the same orientation θ and phase $\varphi + \pi$. We called these lines 'shadow' lines and proposed to eliminate them by a mechanism we called *lateral inhibition*. Roughly speaking, this mechanism acts as follows: the value of a pixel in a cortical image corresponding to orientation θ and phase φ is set to zero, if a pixel of higher value is found in a certain vicinity of the concerned pixel in the cortical image corresponding to orientation θ and phase $\varphi + \pi$. (The shadow lines can also be eliminated by appropriate thresholding as illustrated by the cortical images in the first two columns of Fig. 3.)

Here, it is proposed to extend the lateral inhibition mechanism applying it to all cortical images with the same set of values of the receptive field function parameters, but φ. For Fig. 3 this means that all cortical images in one row interact to generate new such images. The result is shown in Fig. 4. The effect of this mechanism is quite well illustrated by the images in the first row of this figure (Fig. 4(a-d)) where the first two images (a,b) enhance the edges of the body of the bottle and the last image (d) enhances its neck. One has to mention that the practical usefulness of this mechanism for the success of subsequent processing steps has still to be clarified. Another note which should be made is that there is no experimental evidence of the existence of such a mechanism in natural vision system, so that one should consider it as an artificial technique which aims at the simplification of the results with respect to subsequent processing steps.

3. Using cortical images

3.1. Extracting lower dimension representations

Although the sets of cortical images computed according to the above model deliver usefully structured information, they themselves do not give an ultimate solution to the image pattern recognition problem. The achieved structuring does itself not suggest a straightforward scheme for comparing visual input information to prestored information in order to recognise objects. The approach sketched below is a simple straightforward scheme conceived to address this

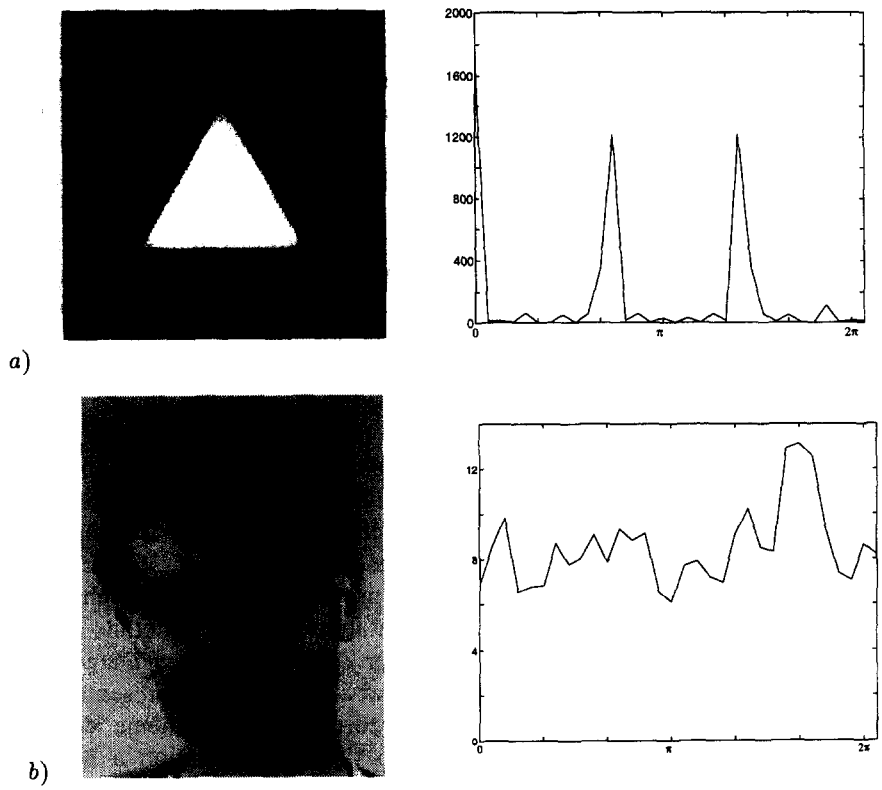

Fig. 5. Two input images and their respective lower dimension representations.

problem. In this approach, developed in cooperation with my students Peter Kruizinga and Tino Lourens and extensively reported elsewhere [17,20–25], all pixel values in a cortical image are summed together to build a quantity which is (partially) characteristic of the input image. (A

Fig. 6. An image A (left) is mapped onto an image B (middle). The image C on the right is obtained by replacing the finest (in this case 8×8) blocks of A by the corresponding best-matching blocks of B. (Image size 500×400.) The energy of the difference of A and C can be used as a measure of dissimilarity. The energy of the divergence of the mapping vector field can be used as a measure of the distortion introduced by the mapping.

positive aspect of this scheme is that the result does not depend on the precise position of the object to be recognised.) Since a number of different cortical channels are used, each of them computing a different cortical image from the input image, a number of such quantities are computed, one per cortical channel. Together these quantities form a descriptor vector which is considered as a projection of the input image onto a point in a lower-dimension space. This representation is then used to discriminate among different images. Fig. 5(a) and Fig. 5(b) show two input images and plots of the respective lower-dimension representations computed by reduction of the cortical images obtained with antisymmetric receptive field functions of a given scale. The energy in a cortical image is plotted as a function of the value of the orientation parameter θ.

The plot in Fig. 5(a) exhibits three dominant maxima which are due to the three edges in the input image. Within the class of convex polygons, this information is sufficient to classify an object with such a lower dimension space representation as a triangle. For a creature which lives in a world of convex polygons, the proposed reduction scheme will solve the classification problem: it only needs an additional stage in its visual system which counts the number of maxima. As to the plot shown in Fig. 5(b), it is hardly possible to infer from its particular structure that it is due to an input image of a human face. Plots which are produced by different face images show a certain degree of similarity (see Fig. 6(a) and Fig. 7 in [17]), but similar plots are produced also by images in which the various face parts do not occupy their proper position. This is due to the fact that this lower dimension representation contains information about the strength and number of edges of particular orientation, but the information about the spatial distribution of these edges is completely lost by applying a global reduction operation.

In order to study the possibilities of the proposed lower-dimension representations, i.e. discrimination between objects within one class, this method was applied to a database of 262 different face images of 36 persons (each person is represented by six to eight face images in the database, technical details can be found in [20]. For each of the face images in the database, a lower-dimension representation was computed according to the scheme sketched above. Based on this representation a nearest-neighbour was searched in the rest of the database. The search was considered to be successful if the nearest neighbour turned out to be the lower-dimension representation of another image of the same person and not successful if it concerned an image of

Fig. 7. The same as in the previous figure but for different persons. Although the blocks used to construct the image C on the right are taken from image B (middle), the result image is similar to image A (left) suggesting that geometrical relations are more important than (exact) illuminance values.

a different person. The search was successful for 256 of the 262 images (97%). This result is astonishing with respect to the simplicity of the method.

3.2. Computing optic flow

In principle, a method which is based on a complete representation has a higher potential than a method based on a representation in a lower-dimension space, because it excludes the possibility of projecting different images onto one and the same point of the representation space. Exploring the possibility to use complete representations, elsewhere [13,14] my student Peter Kruizinga and I proposed to compute a two-dimensional vector field which maps optimally an image onto another image. To compute this vector field, we make use of optic flow techniques applied either to a pair of input images [13] or to a pair of the corresponding sets of cortical images [14]. A mapping is considered to be optimal if it minimises a given dissimilarity measure between the two images, taking into account geometrical constraints such as the relative positions of image blocks. Using the distance achieved for the optical mapping of an image onto another image as a measure for the dissimilarity of the two images, one can search for a nearest neighbour of an input image in a database of prestored images and use the result for recognition.

We make use of a multiscale algorithm to determine the optimal mapping. The method we use is very similar to optic flow algorithms for motion analysis [4]. It involves a number of steps (levels) in each of which one of the levels of the multiscale (pyramid) representation of one of the input images or the set of cortical images computed from it is divided into blocks and for each block a best match is searched in the same level of the corresponding representation of the other image. The algorithm starts at the top of the pyramids and proceeds to their bottoms. At each level, the center of the search space of a block is determined by the result of the search at the previous level.

This algorithm allows to search in a relatively large space using small blocks, without large violations in the relative positions of the blocks. Multiscale resolution beginning with successive refinements of the displacement field is essential, since the direct use of small blocks at the bottom of the pyramids — these can be chosen as small as single pixels — in combination with a large search space may yield a perfect match, i.e. each pixel of an image finds a pixel with the same grayvalue in the other image, but the mapping vector field can be completely random. Fig. 6 and Fig. 7 illustrate the result of the mapping process.

The above described technique was applied to a database of 72 face images of 36 persons, two images per person (this database is a subset of the larger database mentioned in the previous subsection). For each image from the database a dissimilarity measure can be calculated between this image and all other images in the database. Two measures for the dissimilarity of two images A and B were used: (*i*) the energy of the difference image between image A and an image C which is obtained by replacing the finest blocks of A by the corresponding best-matching blocks of image B and (*ii*) the energy of an image which is obtained by computing the divergence of the vector field which maps A onto B. In practice, the first of these quantities proved to give better results than the second or a linear combination of both. The identification was considered correct if the minimum of the dissimilarities of an image of a person to all other images in the database was reached for the other image of the same person. This was the case in 66 of the 72 images (91.6%).

The recognition rate achieved with this method is lower than the recognition rate achieved with the lower-dimension representation method. However, as already mentioned in the previous subsection, the performance of the lower-dimension representation method can easily be degraded. In contrast, the method presented in this subsection does not suffer such deficiencies.

3.3. Using cortical images as input to a self-organising neural network classifier

In contrast to the above methods, the approach presented in this subsection does not make use of an explicit representation of past experiences in the form of a database of prestored

objects, such as the above mentioned descriptor vectors or as sets of input or cortical images. In the method sketched below a self-organising artificial neural network is used which takes as inputs computed cortical images and builds its own internal representations using an unsupervised learning algorithm. When presented a number of input patterns, this system turns out to assign input patterns to clusters of output units according to the classes to which the input patterns belong [26–28].

The system described below was inspired by the fact that in certain hierarchically high areas of the visual cortex of monkeys, more specifically the temporal cortex, cells have been found which react selectively to very complex visual patterns, such as faces [3,19]. These cells are broadly tuned, in that such a cell will respond to images of different faces rather than to images of one particular face. At present there is no computational model of such cells, but the mere fact that they exist is a hint for computational research, in that it is an experimental evidence that the classification of complex visual stimuli proceeds in a convergent way, this means, the presence of a given stimulus is ultimately signaled by individual cells rather than by the collective activation of a population of remote cells [1]. Unfortunately, the facts known about the cortical areas between the primary cortex and the mentioned cortex areas are not sufficient to complete the picture and be able to set up a full computational model allowing automatic classification according to and with the efficacy of the mechanisms employed by natural systems.

Elsewhere [26–28] I proposed to use a self-organising artificial neural network to bridge the gap in our knowledge of the structures which connect orientation-selective cells in the primary visual cortex and class-selective cells in the temporal cortex in order to set up an artificial classification system which exhibits properties similar to those of natural systems. Fig. 8 shows this two-layer artificial neural network constructed of one-dimensional Kohonen networks [12]. In the first layer, there are m such networks, each of n units. Each first-layer network accepts a cortical image as input and the number m of such networks is equal to the number of cortical channels used. A learning sequence consisting of cortical images computed from a sequence of input images is presented to each first-layer network. The result of the learning process is that the units which are activated by images of the same class tend to build clusters. However, clusters in which units are embedded that respond to different classes are observed. Also units are observed which are never activated as well as mixed-class units, i.e. units which react to image patterns that belong to different classes.

The existence of mixed-class units means that, if one would use just one cortical channel with one associated Kohonen network, misclassifications are possible. This is not amazing, since for complex visual input patterns, individual cortical images are not necessarily characteristic of the classes of the corresponding input images. However, *combinations of such images* are characteristic of the class to which an input image belongs.

The two-layer network structure proposed in [26–28] is based on this assumption and expected to function as follows: Each of the first-layer networks associated with the corresponding cortical channels makes its own classification of the cortical image it receives as input. One can think of the first-layer networks associated with different cortical channels as voting for different classes. These votes are counted by another network arranged in the second layer and the class which collects the largest number of votes wins.

The second-layer network has a structure which is similar to the structure of the individual networks in the first layer, in that it is also one-dimensional with the same number of units n and the same neighbourhood relations. The learning

[1] This fact is intriguing in the context of the so-called 'grandmother cell' discussion. Roughly speaking, the content of this discussion is that it is impossible to have one distinct cell for each image which an individual encounters, since each object (or subject, e.g. a grandmother) can have many, actually an infinite number of different views and it would be impossible to assign one distinct cell to each view. This original reasoning is correct, but it is sometimes forgotten and the term 'grandmother cell' is inappropriately used as a synonym of an argument in support of the necessity of distributed representations only.

process for the second layer is started after the learning in the first layer is completed. The learning sequence for the second-layer network consists of a sequence of activity patterns induced in the first-layer networks for the corresponding sequence of input images.

The learning sequence presented to a system consisting of a bank of cortical filters which is followed by the above presented neural network classifier included images of several classes (faces, chairs, bottles, tea cups, triangles) with several images per class. The classification of input patterns presented in the learning sequence was always correct, i.e. images from different classes always activated distinct units. As to test patterns, until now the system was tested only for the classification of images taken under similar conditions as the images used for learning. The system succeeded to classify these test images correctly.

4. Concluding remarks

The computational intensiveness of the computation of cortical images is due to the large number of convolutions (for linear spatial summation) and selection operations (for lateral inhibition) which have to be applied to each input image. Computing one cortical image on a powerful contemporary workstation takes approximately 5 seconds. A set of cortical images can include as many as 320 such images (16 orientations θ, four symmetries φ and five scales σ) which implies that more than half an hour is needed to compute the cortical images set for one single input image. To accelerate computations, the cortical filter bank has been implemented on the Connection Machine CM-5 scale 3 of the University of Groningen (16 nodes, 64 vector units, 2 Gflop/s, 512 Mbyte) and effective acceleration by a factor of nearly forty was achieved. Since the bank consists of a large number of parallel channels and there is a high degree of parallelism in each channel, we are confident that on a more powerful parallel computer — experiments on a Cray J932 (32 processing nodes, 6.4 Gflop/s, 4 Gbyte) to be installed shortly at the University of Groningen are currently being prepared — the cortical filter bank can be realised to give a delay of less than a second per input image which would be comparable with the response times of natural visual systems.

As to the above presented methods for using cortical representations for image recognition and classification, the computational demands are as follows: The time needed for the extraction of a lower-dimension representation is considerably smaller than the time needed to compute the cortical images. As to the database search, the comparison of the lower-dimension representation (320 real numbers) of two images is not

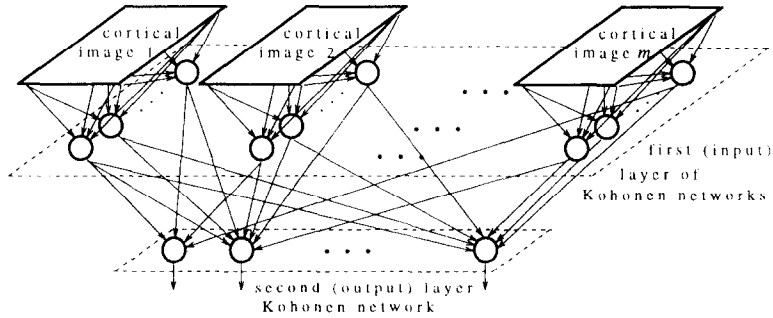

Fig. 8. One one-dimensional Kohonen network is associated with each cortical filter channel, accepting a cortical image as an input. The two-dimensional activity pattern produced by the set of networks in the first layer is used as an input pattern to a second-layer (output) Kohonen network.

computationally demanding but the total search time can become considerable for very large databases. A few hours computing time on the CM5 were needed to collect the identification rate statistics specified in Subsection 3.1.

The computational intensiveness of the second method is due to the matching of blocks in the levels of the multi-scale cortical images representations and is proportional to the size of the input images, the block search space and the number of cortical channels. The problem here is that for each input image pair-wise image comparisons to all images in the database have to be carried out. (This explains the fact that this second method has only been applied to a subset of the database.) A few days computing time on the CM5 were needed to collect the identification rate statistics specified in Subsection 3.2.

The computational intensiveness of the neural network approach presented in Subsection 3.3 depends on the number of units used in the network. The total number of units is proportional to the product of the number of cortical filters used and the number of classes which have to be distinguished. In our first experiments, both of these numbers were relatively small (32 cortical channels and 5 image classes), so that computing time did not present a major problem both for learning and classification. The real time learning and classification for a large number of (320) cortical channels and thousands of different image classes exceed the power of the currently available parallel supercomputers.

As demonstrated by the above examples, studying and simulating the principles and mechanisms employed by natural vision systems can lead to new image analysis and object recognition techniques which have the potential to outperform traditional machine vision approaches. Our further research will focus on studying and applying the mechanisms used by natural vision systems to extract information from cortical representations. Another topic will be the extension of the above presented simple cell primary cortex representations with representations which correspond to the responses of so-called complex cells and grating cells which do not abide to the simple cell model [15].

Acknowledgements

The investments in the Connection Machine CM-5 of the University of Groningen were partly supported by the Foundation for Computer Science Research (SION) and the Organisation for Scientific Research (NWO) of The Netherlands.

References

[1] J.G. Daugman, Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *J. Optical Soc. Amer. A* 2 (7) (1985) 1160–1169.

[2] J.G. Daugman, Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression, *IEEE Trans. Acoustics, Speech and Signal Processing* 36 (7) (1988) 1169–1179.

[3] C. Gross, Representation of visual stimuli in inferior temporal cortex, *Phil. Trans. Roy. Soc. Lond. B: Biolog. Sci.* 335 (1992) 3–10.

[4] D.J. Heeger, Model for the extraction of image flow, *J. Opt. Soc. Amer. A* 3 (8) (1987) 1455–1471.

[5] F. Heitger, L. Rosenthaler, R. Von der Heydt, E. Peterhans and O. Kübler, Simulation of neural contour mechanisms: from simple to end-stopped cells, *Vision Res.* 23 (5) (1992) 963–981.

[6] D.H. Hubel and T. Wiesel, Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex, *J. Physiol.* (London) 160 (1962) 106–154.

[7] D.H. Hubel, Explorations of the primary visual cortex, 1955–1978" (1981 Nobel Price lecture), *Nature* 299 (1982) 515–524.

[8] E. Kaplan and R.M. Shapley, The primate retina contains two types of ganglion cells, with high and low contrast sensitivity, *Proc. Natl. Acad. Sci. USA* 83 (1986) 2755–2757.

[9] S.W. Kuffler, Discharge patterns and functional organization of mammalian retina, *J. Neurophysiol.* 16 (1953) 37–68.

[10] *Grand Challenges 1993: High Performance Computing and Communications*, The FY 1993 U.S. Research and Development Program (Report by the Committee on Physical, Mathematical and Engineering Sciences) Federal Coordinating Council for Science, Engineering and Technology, Office of Science and Technology Policy, 1993 (68 p.).

[11] J.P. Jones and L.A. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, *J. Neurophysiol.* 58 (1987) 1233–1258.

[12] T. Kohonen, *Self Organization and Associative Memory* (Berlin, Springer-Verlag, 1990, third ed.).

[13] P. Kruizinga and N. Petkov, Optic flow applied to person identification, in J.C. Zuidervaart and L. Dekker,

eds., *Proc. 1994 EUROSIM Conf. on Massively Parallel Processing*, (June 21–23, 1994) Delft, The Netherlands (Amsterdam, Elsevier, 1994) 871–878.

[14] P. Kruizinga and N. Petkov, Face recognition and person identification based on multiscale matching of cortical images, Proc. Int. Conference on High-Performance Computing and Networking, HPCN Europe '95, May 2–5, 1995, Milano, Italy. B. Hertzberger and G. Serazzi, eds., *Lecture Notes on Computer Science* 919 (Berlin, Springer-Verlag, 1995) 420–427.

[15] P. Kruizinga and N. Petkov, A computational model for periodic-pattern-selective cells, Proc. Int. Workshop on Artificial Neural Networks, IWANN'95, June 7–9, 1995, Torremolinos (Málaga), Spain. J. Mira and P. Sandoval, eds., *Lecture Notes on Computer Science* 930 (Berlin, Springer-Verlag, 1995) 90–99.

[16] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. Von der Malsburg, R.P. Würtz and W. Konen, Distortion invariant object recognition in the dynamic link architecture, *IEEE Trans. Comput.* 42 (3) (1993) 300–311.

[17] T. Lourens, N. Petkov and P. Kruizinga, Large scale natural vision simulations, *Future Generation Comput. Syst.* 10 (1994) 351–358.

[18] B.S. Manjunath and R. Chellappa, A unified approach to boundary perception: edges, textures, and illusory contours, *IEEE Trans. Neural Networks* 4 (3) (1993) 96–108.

[19] D.I. Perrett, J.K. Hietanen, M.W. Oram and P.J.Benson, Organization and functions of cells responsive to faces in the temporal cortex, *Phil. Trans. Roy. Soc. Lond. B: Biolog. Sci.* 335 (1992) 23–30.

[20] N. Petkov, P. Kruizinga and T. Lourens, Biologically motivated approach to face recognition, *Proc. Int. Workshop on Artificial Neural Networks* (June 9–11, 1993), Sitges (Barcelona), Spain (Berlin, Springer-Verlag, 1993) 68–77.

[21] N. Petkov, P. Kruizinga and T. Lourens, Lateral inhibition in cortical filters, *Proc. Int. Conf. on Digital Signal Processing and Int. Conf. on Computer Applications to Engineering Systems* (July 14–16, 1993), Nicosia, Cyprus, pp. 122–129.

[22] N. Petkov, P. Kruizinga and T. Lourens, Natural visual simulations -- An application to face recognition, in H. Dedieu, ed., *Proc. 1993 Europ. Conf. on Circuit Theory and Design*, Aug. 30–Sept. 3, 1993, Davos, Switzerland (Amsterdam, Elsevier Sci. Publ., 1993) 821–826.

[23] N. Petkov, P. Kruizinga and T. Lourens, Face recognition on the Connection Machine CM-5", in G.R. Joubert, D. Trystram, F.J. Peters and D.J. Evans, eds.,: *Parallel Computing: Trends and Applications, Proc. Int. Conf. ParCo 93*, Sept. 7–10, 1993, Grenoble, France, *Advances in Parallel Computing*, Vol.9 (Amsterdam, Elsevier Sci. Publ., 1994) 185–192.

[24] N. Petkov, P. Kruizinga and T. Lourens, Orientation competition in cortical filters — An application to face recognition, *Computing Science in The Netherlands 1993*, Nov. 9–10, 1993, Utrecht (Amsterdam, Stichting Mathematisch Centrum, 1993) 285–296.

[25] N. Petkov and T. Lourens, Interacting cortical filters for object recognition, *Proc. Asian Conf. on Computer Vision*, Nov. 23–25, 1993, Osaka, Japan, pp. 583–586.

[26] N. Petkov, Image classification system inspired by neurophysiological findings, in D. Docampo and A.R. Figueiras, eds., *Proc. Final COST 229 Workshop Adaptive Systems, Intelligent Approaches, Massively Parallel Computing and Emergent Techniques in Signal Processing and Communications*, Bayona (Vigo), Spain (Oct. 1994) (Vigo, University of Vigo, 1994) 209–213.

[27] N. Petkov, Cortical images, self-organising neural networks and object recognition, in M. Suk (ed.) *Shape Analysis* (Ablex Publishing, 1995, in print) 40 p.

[28] N. Petkov, *Biologically motivated image classification system*, in ed. Ph. Laplante and A. Stoyenko, *Real-Time Imaging* (Academic Press, 1995, in print) 31 p.

N. Petkov holds a chair of Parallel Computing at the University of Groningen. He is also head of the Centre for High Performance Computing (HPC) of the same university. Dr. Petkov is the author of two books and of further 60 scientific publications in the area of parallel processing and holds several patents. His current research interests are in the area of Vision and Cognition as Grand Challenge problems, in particular in conceiving computationally very intensive biologically motivated methods.